EXHIBIT 1

"

"

The Current Population Survey Voting and Registration Supplement Overstates Minority Turnout

Stephen Ansolabehere, Harvard University
Bernard L. Fraga, Emory University
Brian F. Schaffner, Tufts University

The Current Population Survey (CPS) is a key source of information on who votes. Turnout estimates derived from the CPS are often cited in academic research on participation, widely used in the calibration of surveys, and central to ongoing legal and policy debates over the protection of voting rights in the United States. We compare CPS estimates to official voter turnout records from 2008–18 and document consistent, significant discrepancies that call into question the reliability of CPS turnout statistics. Specifically, the CPS overestimates black and Hispanic turnout relative to non-Hispanic whites, whether relying on turnout rates as a share of eligible citizens or the racial/ethnic composition of the voting population. Sampling error and commonly used adjustments to CPS estimates do not account for or correct this bias, and thus academics and policy makers should use discretion when judging recent shifts in voter turnout with survey data.

Researchers have long relied on the Current Population Survey's (CPS) Voting and Registration Supplement as an authoritative source of data on turnout in the United States, including in studies examining turnout among racial and ethnic minority groups (Francia and Orr 2014; Jackson 2003; Rocha et al. 2010; Sides, Schickler, and Citrin 2008; Wolfinger, Highton, and Mullin 2005). For example, Leighley and Nagler use the CPS to conclude that "where thirty years ago black turnout lagged substantially behind white turnout, in 2008 it was the same as white turnout" (2013, 31). Think tanks and policy makers rely heavily on the CPS to craft public policies governing elections. And in 2013, Chief Justice John Roberts pointed to CPS data as central to the court's decision to strike down Section 4(b) of the Voting Rights Act in *Shelby County v. Holder*. Using CPS estimates, the court concluded "that African-American voter turnout exceeded white voter turnout in five of the six States originally covered." The court's reading of the CPS data was correct, but

there are growing concerns among researchers that the CPS may not accurately measure turnout, particularly for minority citizens (Bauman 2018; McDonald 2007; McKee, Hood, and Hill 2012).

The CPS data are also used for calibrating weights for surveys that are widely used to understand American politics. The Cooperative Congressional Election Study (CCES), exit polls, Associated Press's VoteCast, and many other surveys all use the CPS to calibrate weights for voters or registered voters. The CPS is also used in recent applications of multilevel regression and poststratification (MRP), both in political science as well as election polling and modeling (Ghitza and Gelman 2013).

While the government-produced CPS performs better than many other surveys in collecting data on political participation, it is still susceptible to problems such as misreporting and nonresponse (Ansolabehere and Hersh 2012; Bernstein, Chadha, and Montjoy 2001; Cassel 2004; Cuevas-Molina

2017; Enamorado and Imai 2018; Jackman and Spahn 2019; Traugott and Katosh 1979). The CPS relies on respondents to self-report whether they are registered to vote and whether they voted in the most recent election. If some groups have more errors in self-reported turnout than others, it may lead to inaccurate conclusions when comparing participation rates (Cuevas-Molina 2017; Enamorado and Imai 2018). For example, McKee et al. (2012) find that black turnout rates in Georgia in 2008 were significantly lower than what the CPS reported.

Does the CPS provide an accurate picture of turnout patterns across racial and ethnic groups? We test the validity of the CPS's turnout estimates by comparing those rates to what voter files indicate. We show that the CPS overstates turnout rates more for minority voters than for whites and that this bias cannot be easily corrected. Our findings thus raise concerns about the use of the CPS for statistical analysis of participation.

COMPARING VOTER FILE AND CPS TURNOUT ESTIMATES

We compare turnout for different racial groups as estimated by the CPS with turnout for those same groups based on voter files. A growing body of academic literature relies on voter files to study electoral behavior, and voter records are now a key source of information used to validate survey self-reports of turnout (Ansolabehere and Hersh 2012; McDonald 2007). We rely on data from six southern states where self-reported race/ethnicity is provided during the voter registration process: Alabama, Florida, Georgia, Louisiana, North Carolina, and South Carolina.[1] For these states we can determine how accurately racial/ethnic differences in voter turnout are reflected in the CPS.

The voter file data were provided by the voter file firm Catalist. Catalist has an established track record of producing estimates of the total number of voters that closely align with semiofficial statistics as computed by McDonald (2017; see also Fraga 2018). The voter file data are the baseline comparison, but they themselves may have flaws (Nyhan, Skovron, and Titiunik 2017). If vote history was incomplete in some counties or states due to faulty record keeping, then the voter file data would not be a valid point of comparison. However, we find that the voter file data provided by Catalist do not suffer from any widespread lack of coverage. In fact, as we show in the appendix, the turnout estimates using vote history (from Catalist) and official turnout figures differ by no more than 1.3% from 2008 to 2018, and these deviations do not

appear to be more common in precincts with higher numbers of minority voters.

Turnout rate calculations consist of a numerator, the number of people of a given race or ethnicity who voted, and a denominator, the number of people of a given race or ethnicity who were eligible to vote. We gathered the number of individuals who are listed as having voted in general elections from 2008 to 2018 by self-reported race/ethnicity, establishing these figures to be the most valid available count of the number of persons voting by race in each of the six states. We estimate the eligible electorate for each racial group using US Census Bureau Population Estimates Program and American Community Survey data. As discussed in appendix section A.1, we construct a citizen voting-age nonincarcerated population estimate that approximates the denominator used in CPS calculations.

We compare voter file and census derived statistics to the CPS estimates in the same six states. Drawing on the publicly available, raw, individual-level November CPS Voting and Registration Supplement data, we are able to precisely replicate the number of voters and rates of voter turnout reported by the Census Bureau in their official publications. The CPS also calculates turnout rates by dividing their tally of the number of persons who voted by the sampled population that is both over 18 and is a naturalized or native-born US citizen.[2]

The CPS estimate of turnout is based on a sample, and thus random sampling error may account for deviations we see from the true population parameter. The CPS uses a multiwave household-based stratified sampling scheme in order to obtain data on over 140,000 Americans every month. They do not provide data on the sampling units they use but do provide margins of error based on generalized variance parameters for some top-line statistics. Previous research establishes that inferences relying on CPS-provided measures of uncertainty are unreliable (Davern et al. 2006). Following recommendations from Davern et al. (2007), we instead use the available geographic and household indicators to approximate this complex sampling scheme and compute 95% confidence intervals for CPS rates of voter turnout by state and race/ethnicity.

RESULTS

Figure 1 shows the overall turnout rate in elections held from 2008 through 2018 as well as the CPS estimates for those rates. Figure 1A presents these comparisons nationally. The

---

1. Mississippi and Tennessee also ask registrants to volunteer their race/ethnicity, but in the above-mentioned states well over 90% of registrants provide this information.

2. Those not responding to CPS voting questions are considered nonvoters, a practice questioned by many researchers. We use CPS reported figures for consistency with official estimates and discuss the impact of nonresponse on CPS estimation error later in the article.

Figure 1. Overall voter turnout, Current Population Survey (CPS) versus state estimates: *A*, national; *B*, six southern states. Black points and confidence intervals indicate estimates from the CPS. Gray points are estimates of the number of voters in each election divided by the nonimprisoned citizen voting-age population, as estimated by McDonald (2017).

CPS estimates are presented with 95% confidence intervals, although these are smaller than the dots in the plot. In presidential election years, the CPS overestimates the turnout rate by between .4 and 2.4 percentage points. In midterm election years, when turnout is lower, the CPS overestimates turnout by 2.7–4.7 percentage points. Only in 2008 does the actual turnout rate fall within the 95% confidence interval for the CPS estimate. Generally speaking, the discrepancies in figure 1 grow larger when turnout drops, suggesting that turnout self-reports captured by the CPS may fluctuate less than turnout itself.

Figure 1*B* shows the same overall turnout comparisons, but this time limiting the comparison to the six southern states that we focus on for our comparison of turnout rates by race/ethnicity. The overall patterns are similar; however, in the most recent two elections, the actual turnout rate falls within the 95% confidence interval of the CPS estimate.

Now we turn to our comparison of actual turnout rates and CPS estimates for the non-Hispanic white, African American, and Hispanic populations in the six southern states. These are presented in figure 2. The first important point from this figure is that black turnout did not, in fact, exceed white turnout in 2012. In every election plotted in the figure, the turnout rate among African Americans based on the voter file data lags behind the turnout rate among whites. In two recent elections,

this gap has been quite large. In 2016, the turnout rate among whites was 10.8 percentage points higher than the turnout rate among blacks. In 2018, the gap in turnout rates was 8.6 percentage points. Yet, researchers relying on the CPS to examine turnout rates in these six states would find a gap in participation rates of just 3.7 points in 2016 and 2.5 points in 2018.

The plots in figure 2 document why this happens—the CPS consistently overstates the participation rate among blacks and Hispanics, while it sometimes underestimates participation among whites. Figure 2*A* shows that in two of the three presidential elections, the CPS accurately estimates
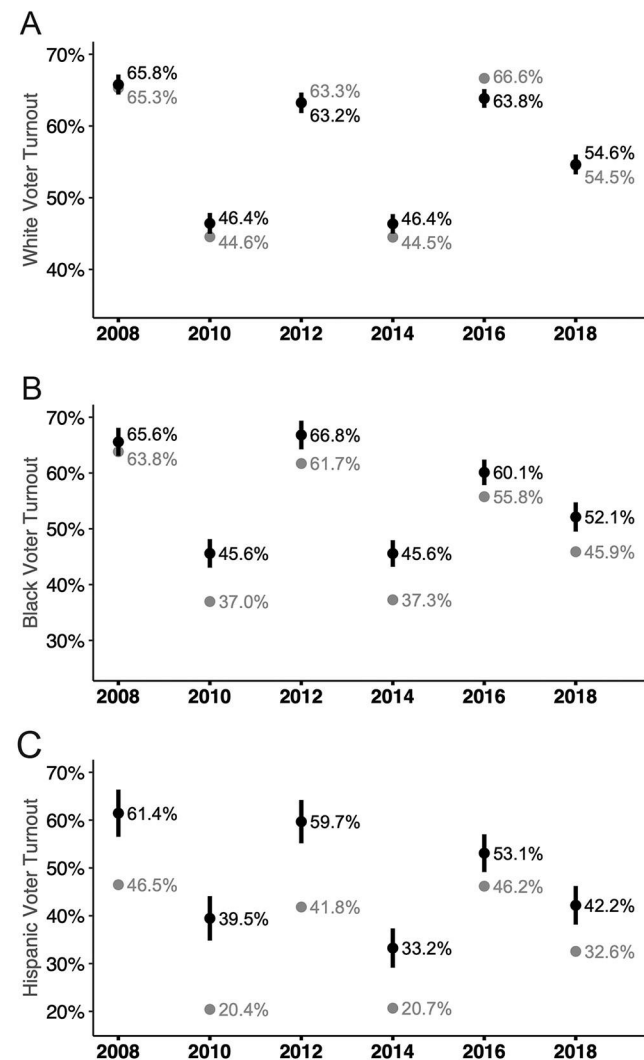

Figure 2. Voter turnout by race, Current Population Survey (CPS) versus voter file estimates: *A*, white voter turnout; *B*, black voter turnout; *C*, Hispanic voter turnout. Includes voters in six southern states with comprehensive data on race/ethnicity of registrants (AL, FL, GA, LA, NC, SC). Black points and 95% confidence intervals indicate estimates from the CPS. Gray points are estimates of the number of voters by race/ethnicity in each election divided by the nonimprisoned citizen voting-age population by race, using data from Catalist and the US Census Bureau.

the turnout rate among whites in these six states, although in 2016 the CPS data actually underestimated white turnout by 2.8 points. In midterm elections held in 2010 and 2014, the CPS data overstated the white turnout rate by about 2 points.

Figure 2B shows that the CPS overstated the turnout rate among blacks by more than 4 percentage points in five of the six elections. Even in 2008, the peak year for African American turnout, the CPS still overestimated black turnout by nearly 2 percentage points (although it was within the 95% confidence interval). Figure 2C shows a similar pattern for Hispanics. The CPS estimates for Hispanic turnout are always much higher than the figure suggested by the voter files.

While turnout rates are often the key statistic for those studying turnout, there is sometimes more interest in the racial composition of those who voted, particularly among analysts who use the CPS to calibrate survey weights or MRP estimates (e.g., Ghitza and Gelman 2013). On this metric, the story is much the same. Figure 3A shows that in the six southern states, the CPS almost always produces an estimate of the white share of the electorate that is lower than what it actually was, although on several occasions the actual composition was within the 95% confidence interval. By contrast, figure 3B shows that in four of the six elections, the CPS estimate of the black share of the electorate is statistically significantly higher than the actual percentage of voters who were black. Finally, figure 3C shows that the CPS overstates the Hispanic share of the electorate by a statistically significant amount in all but one election cycle. The CPS appears to present a picture of the electorate that is more racially diverse than what the voter files indicate.

WHY DO THE VOTER FILE AND CPS ESTIMATES DIFFER?

There are at least four reasons why the CPS and vote history data may differ. First, nonresponse to questions about registration and turnout in the CPS may cause overestimates of turnout. Nonresponse to the turnout section of the CPS survey varies by the race/ethnicity of the respondent (Fraga 2018), which could cause the bias in these estimates if all nonrespondents are assumed to be nonvoters. Hur and Achen (2013) suggest a series of adjustments to the CPS data that instead drop those who do not answer the registration and turnout questions, then reweight state estimates to McDonald's (2017) turnout data. This technique assumes that, conditional on responding to the CPS voting questions, subgroup turnout rates are accurately measured. However, as we show in the appendix, applying these corrections produces estimates of white voter turnout that are biased even further downward in the six states we examine and continues to produce estimates of black and Hispanic turnout that are as much as 13 per-
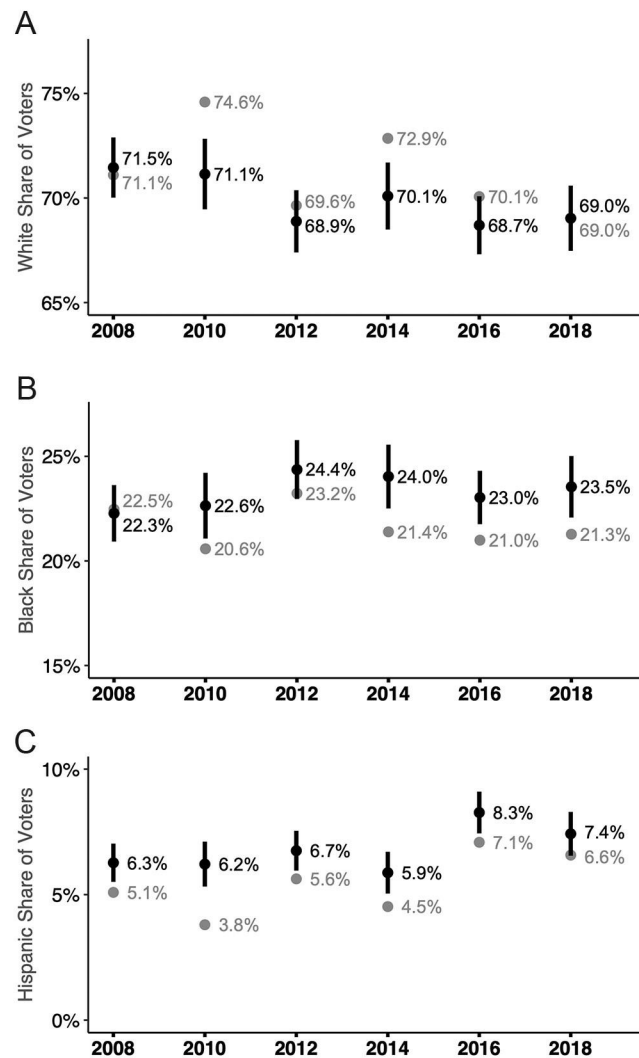

Figure 3. Share of voters by race, Current Population Survey (CPS) versus voter file estimates: A, white share of voters; B, black share of voters; C, Hispanic share of voters. Includes voters in six southern states with comprehensive data on race/ethnicity of registrants (AL, FL, GA, LA, NC, SC). Black points and 95% confidence intervals indicate estimates from the CPS. Gray points are estimates of the share of voters by race/ethnicity in each election, using data from Catalist.

centage points greater than what the voter file indicates. Shares of the electorate (shown in fig. 3), which are not denominator dependent, are also biased in favor of whites and, in the case of black and Hispanic shares, well outside of the margin of error in recent elections.

A second possibility would be incomplete record keeping on the state voter files, particularly if there were more missing data among minorities than for whites (Nyhan et al. 2017). If voter files were significantly undercounting minority turnout, such a pattern would be evident from the fact that the turnout records in the vote history files would not match the official certified vote counts. In table A.1 (tables A.1–A.5 are available online), we show that the voter file counts are very close to the

official vote tallies provided by the states. Additionally, when we look at calculations related to each group's share of the electorate (fig. 3) the discrepancies we document persist. Therefore, it is unlikely that missing data on the voter files can account for these turnout differences.

A third possibility is that minority voters might misreport turnout at higher rates than whites. Voter validation studies generally find higher rates of overreporting turnout among minorities (Enamorado and Imai 2018; McKee et al. 2012).[3] In table A.3, we show how overreporting is more common among black and Hispanic respondents in the CCES surveys. This provides some support for the misreporting explanation, although we do not know whether these patterns would also exist in the CPS survey. Misreporting may be less pronounced in the CPS survey since it is not inherently a political survey. To fully test this hypothesis, it would be necessary to match CPS respondents to voter file data to determine the extent to which overreporting might be causing errors.

A fourth possible explanation for the patterns we observe may relate to nonresponse bias or differential panel attrition. Individuals recruited for the CPS take surveys in four consecutive months and then are recontacted eight months later to take four more surveys in consecutive months. Participation in the CPS is voluntary, and panel attrition is significant; one estimate is that panel attrition is 20% over the year between the first and the fifth wave (Rivera Drew, Flood, and Warren 2014). Studies of panel attrition in other surveys have found that nonvoters are more likely to attrit than voters and nonwhites also tend to have higher attrition rates (Smith and Son 2010). If black and Hispanic nonvoters are especially likely to attrit from the CPS, then it may lead the sample to have a higher share of black and Hispanic voters relative to the population. Since no CPS panel spans two election cycles, it is not possible to test this hypothesis without matching the CPS sample to voter file data.

CONCLUSION

The analysis presented here points to a serious bias in the CPS turnout estimates, one that has significant consequences for the inferences we make about racial/ethnic disparities in turnout rates. The CPS relies on a standard approach to studying turnout—simply ask survey respondents whether they voted. But biases in the composition of the CPS samples or in responses to the turnout questions are likely responsible for the errors we document here (Dahlgaard et al. 2019; Lahtinen et al. 2019; Smith and Son 2010). To better diagnose

the problem, the CPS should conduct a voter validation study akin to those undertaken by other surveys. A validation of the CPS Voting and Registration Supplement would help to document the biases in the estimates produced by the Census Bureau and may open the door to methodological innovations that ensure policy debates are informed by data accurately reflecting the composition of the American electorate. In the meantime, we suggest that analysts use caution when making inferences about variation in turnout rates by racial and ethnic groups.

REFERENCES

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." *Political Analysis* 20 (4): 437–59.

Bauman, Kurt. 2018. "How Well Does the Current Population Survey Measure the Composition of the U.S. Voting Population?" Working paper no. 2018-25, US Census Bureau, Washington, DC. https://www.census.gov/library/working-papers/2018/demo/SEHSD-WP2018-25.html.

Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. "Overreporting Voting: Why It Happens and Why It Matters." *Public Opinion Quarterly* 65 (1): 22–44.

Cassel, Carol A. 2004. "Voting Records and Validated Voting Studies." *Public Opinion Quarterly* 68:102–8.

Cuevas-Molina, Ivelisse. 2017. "Voter Turnout Overreports: Measurement, Modeling and Deception." PhD diss., University of Massachusetts Amherst.

Dahlgaard, Jens Olav, Jonas Hedegaard Hansen, Kasper M. Hansen, and Yosef Bhatti. 2019. "Bias in Self-Reported Voting and How It Distorts Turnout Models: Disentangling Nonresponse Bias and Overreporting among Danish Voters." *Political Analysis* 27 (4): 590–98.

Davern, Michael, Arthur Jones Jr., James Lepkowski, Gestur Davidson, and Lynn A. Blewett. 2006. "Unstable Inferences? An Examination of Complex Survey Sample Design Adjustments Using the Current Population Survey for Heath Services Research." *Inquiry* 43 (3): 283–97.

Davern, Michael, Arthur Jones Jr., James Lepkowski, Gestur Davidson, and Lynn A. Blewett. 2007. "Estimating Regression Standard Errors with Data from the Current Population Survey's Public Use File." *Inquiry* 44 (2): 211–24.

Enamorado, Ted, and Kosuke Imai. 2018. "Validating Self-Reported Turnout by Linking Public Opinion Surveys with Administrative Records." SSRN. https://doi.org/10.2139/ssrn.3217884.

Fraga, Bernard L. 2018. *The Turnout Gap: Race, Ethnicity, and Political Inequality in a Diversifying America*. New York: Cambridge University Press.

Francia, Peter L., and Susan Orr. 2014. "Labor Unions and the Mobilization of Latino Voters: Can the Dinosaur Awaken the Sleeping Giant?" *Political Research Quarterly* 67 (4): 943–56.

Ghitza, Yair, and Andrew Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns among Small Electoral Subgroups." *American Journal of Political Science* 57 (3): 762–76.

Hur, Aram, and Christopher H. Achen. 2013. "Coding Voter Turnout Responses in the Current Population Survey." *Public Opinion Quarterly* 77 (4): 985–93.

Jackman, Simon, and Bradley Spahn. 2019. "Why Does the American National Election Study Overestimate Voter Turnout?" *Political Analysis* 27 (2): 193–207.

Jackson, Robert A. 2003. "Differential Influences on Latino Electoral Participation." *Political Behavior* 25 (4): 339–66.

---

3. Some studies find that racial differences in overreporting disappear after controlling for other factors, but what matters here is whether such patterns manifest in the aggregate.

Lahtinen, Hannu, Pekka Martikainen, Mikko Mattila, Hanna Wass, and Lauri Rapeli. 2019. "Do Surveys Overestimate or Underestimate Socioeconomic Differences in Voter Turnout? Evidence from Administrative Registers." *Public Opinion Quarterly* 83 (2): 363–85.

Leighley, Jan E., and Jonathan Nagler. 2013. *Who Votes Now? Demographics, Issues, Inequality, and Turnout in the United States*. Princeton, NJ: Princeton University Press.

McDonald, Michael P. 2007. "The True Electorate: A Cross-Validation of Voter Registration Files and Election Survey Demographics." *Public Opinion Quarterly* 71 (4): 588–602.

McDonald, Michael P. 2017. "Voter Turnout." United States Elections Project. http://www.electproject.org/home/voter-turnout/voter-turnout-data.

McKee, Seth C., M. V. Hood III, and David Hill. 2012. "Achieving Validation: Barack Obama and Black Turnout in 2008." *State Politics and Policy Quarterly* 12 (1): 3–22.

Nyhan, Brendan, Christopher Skovron, and Rocío Titiunik. 2017. "Differential Registration Bias in Voter File Data: A Sensitivity Analysis Approach." *American Journal of Political Science* 61 (3): 744–60.

Rivera Drew, Julia A., Sarah Flood, and John Robert Warren. 2014. "Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records across 16 Months." *Journal of Economic and Social Measurement* 39 (3): 121–44.

Rocha, Rene R., Caroline J. Tolbert, Daniel C. Bowen, and Christopher J. Clark. 2010. "Race and Turnout: Does Descriptive Representation in State Legislatures Increase Minority Voting?" *Political Research Quarterly* 63 (4): 890–907.

Sides, John, Eric Schickler, and Jack Citrin. 2008. "If Everyone Had Voted, Would Bubba and Dubya Have Won?" *Presidential Studies Quarterly* 38 (3): 521–39.

Smith, Tom William, and Jaesok Son. 2010. "An Analysis of Panel Attrition and Panel Change on the 2006–2008 General Social Survey Panel." GSS Methodological report no. 118, NORC, University of Chicago.

Traugott, Michael W., and John P. Katosh. 1979. "Response Validity in Surveys of Voting Behavior." *Public Opinion Quarterly* 43:359–77.

Wolfinger, Raymond E., Benjamin Highton, and Megan Mullin. 2005. "How Postregistration Laws Affect the Turnout of Citizens Registered to Vote." *State Politics and Policy Quarterly* 5 (1): 1–23.